

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11) EP 0 768 381 A2

(12)

## EUROPEAN PATENT APPLICATION

(43) Date of publication:  
16.04.1997 Bulletin 1997/16

(51) Int. Cl.<sup>6</sup>: C12N 15/52, C12N 9/00,  
C12N 15/11, C12N 15/70,  
C12N 1/21

(21) Application number: 96116005.8

(22) Date of filing: 07.10.1996

(84) Designated Contracting States:  
DE FR GB

(30) Priority: 09.10.1995 JP 261147/95

(71) Applicant: THE INSTITUTE OF PHYSICAL &  
CHEMICAL RESEARCH  
Wako-shi Saitama 351-01 (JP)

(72) Inventors:  
• Kamiya, Yuji  
c/o The Institute of Phys. and Chem.  
Saitama 351-01 (JP)  
• Yamaguchi, Shinjiro  
Suginami-ku, Tokyo 166 (JP)

(74) Representative: Godemeyer, Thomas, Dr.  
Hauptstrasse 58  
51491 Overath (DE)

### (54) Kaurene synthase

(57) A polypeptide characterized by amino acid Nos. 1-789 of the amino acid sequence defined by SEQ ID NO:1 of the sequence listing or a polypeptide having the said amino acid sequence incorporated by insertion, deletion, or substitution of one or more amino acid residues. The peptides have catalytic activities of converting copalyl pyrophosphate to ent-kaurene and are useful for the regulation of plant growth. Also provided are DNAs encoding the polypeptides.

EP 0 768 381 A2

### Description

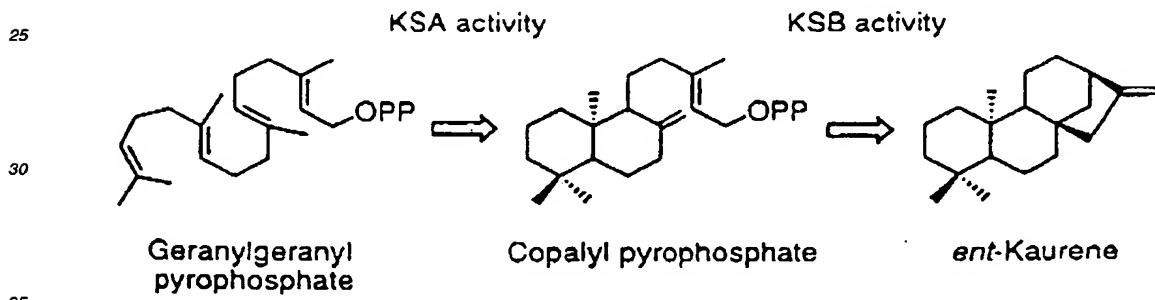
## BACKGROUND OF THE INVENTION

The present invention relates to a novel kaurene synthase and a DNA which encodes the enzyme.

ent-Kaurene is an important intermediate in gibberellin (GA) biosynthesis, and is synthesized from geranylgeranyl pyrophosphate (GGPP) via copalyl pyrophosphate (CPP). These steps are catalyzed by ent-kaurene synthase A (KSA) and ent-kaurene synthase B (KSB), respectively (Coolbaugh, R.C., *The Biochemistry and Physiology of Gibberellins*, Vol. 1, Pracger Publishers, New York, pp.53-98, 1983). Other terpenoids such as carotenoids, xanthophyll, and phytol are also synthesized from geranylgeranyl pyrophosphate as a precursor, while copalyl pyrophosphate is used as precursors for macrocyclic diterpene biosyntheses. ent-kaurene synthase A and ent-kaurene synthase B are thus important enzymes in the initial stage of gibberellin biosynthesis (Coolbaugh, *supra*; Chung C.H. et al., *Plant Physiol.*, 80, pp.544-548, 1986; Graebe, J.E., *Annu. Rev. Plant. Physiol.*, 38, pp.419-465, 1987).

Duncan et al. separated ent-kaurene synthase A and ent-kaurene synthase B from native cucumber (*Marah macrocarpus* L.) using a QAE column. The authors suggested that the conversion of geranylgeranyl pyrophosphate to ent-kaurene is catalyzed by the two distinguishable enzymes (Duncan, J.D. and West C.A., *Plant Physiol.*, 68, pp.1128-1134, 1981). In this article, it is shown that ent-kaurene synthase A and ent-kaurene synthase B closely associated with each other during the ent-kaurene synthesis and that ent-kaurene synthase B preferentially utilized endogenous copalyl pyrophosphate produced by ent-kaurene synthase A rather than exogenous copalyl pyrophosphate.

As shown in the scheme set out below, ent-kaurene synthase A catalyzes the conversion of geranylgeranyl pyrophosphate (GGPP) to copalyl pyrophosphate (CPP), and the resulting copalyl pyrophosphate is successively converted to ent-kaurene by ent-kaurene synthase B.



In recent years, cDNA clonings of gibberellin biosynthases have been reported. GAI locus of *Arabidopsis* was isolated by genomic subtraction (Sun, T.P., et al., *Plant Cell*, 4, pp.119-128, 1992) and the corresponding cDNA clone was verified to encode ent-kaurene synthase A by over-expression system in *E. coli* (Sun T.P. and Kamiya, Y., *Plant Cell*, 6, pp.1509-1518, 1994). *Anl* locus of maize was cloned by the transposon tagging technique and the deduced amino acid sequence from the cDNA was revealed to have significant homology with *Arabidopsis* ent-kaurene synthase A, although the function of the expressed protein remains unknown (Bensen, R.J., et al., *Plant Cell*, 7, pp.75-84, 1995). A putative 3b-hydroxylase in gibberellin biosynthesis was also cloned by the T-DNA tagging method using *Arabidopsis* GA4 mutant (Chiang, H.-h., et al., *Plant Cell*, 7, pp.195-201, 1995).

As for ent-kaurene synthase B, the enzyme was partially purified from *Ricinus communis* (Spickett, C.M., et al., *Phytochem.*, 37, pp.971-973, 1994). *Ricinus communis* biosynthesizes kaurene and its family diterpenes, i.e., bayarene, torakiroban, and sandarakopymaradiene. These compounds are synthesized through selective cyclizations from a common intermediate, copalyl pyrophosphate. Although a cloning of an ent-kaurene synthase B gene was attempted (15th International Conference on Plant Growth Substance, July 14-18, 1995, Minneapolis, USA, Kamiya, Y., et al., Subject No. 040 and Saito, T., et al., Subject No. 110), the complete sequence of a cDNA and the amino acid sequence of the enzyme has not been reported.

## SUMMARY OF THE INVENTION

55 An object of the present invention is to provide a purified ent-kaurene synthase B. Another object of the present invention is to provide the amino acid sequence of the enzyme and a DNA encoding the enzyme.

The present inventors succeeded in purifying an ent-kaurene synthase B from endosperm of pumpkin (*Cucurbita maxima* L.) seeds. The inventors revealed its amino acid sequence (a partial sequence), and based on the information,

of the amino acid sequence, they succeeded in cloning a gene which encodes ent-kaurene synthase B.

Accordingly, the present invention provides a polypeptide which is characterized by the amino acid sequence as defined by SEQ ID NO:1 of the sequence listing (amino acids No. 1-789; the amino acid sequence is shown by the IUPAC one-letter symbols). The polypeptide can act as an enzyme having the catalytic activity of converting copalyl pyrophosphate to ent-kaurene (in the specification, this enzyme will be occasionally referred to as ent-kaurene synthase B, or more simply as KSB). According to another embodiment of the present invention, DNAs encoding the above-described polypeptide are provided. As a preferred embodiment thereof, a DNA which is characterized by nucleotides Nos. 138-2507 (including the termination codon) of the nucleic acid sequence as defined by SEQ ID NO:2 of the sequence listing is also provided.

10

#### BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a chromatogram profile of the ent-kaurene synthase B of the present invention applied to hydrophobic interaction chromatography.

15

Fig. 2 is a chromatogram profile of the ent-kaurene synthase B of the present invention purified by a DEAE column.

Fig. 3 is a chromatogram of an ent-kaurene synthase B of the present invention purified by a hydroxyapatite column.

Fig. 4 shows activities of an ent-kaurene synthase B of the present invention in the presence of divalent cations.

20

#### DETAILED EXPLANATION OF THE INVENTION

In addition to the aforementioned naturally derived ent-kaurene synthase B from pumpkin, any polypeptides also fall within the scope of the polypeptides provided by the present invention which have the amino acid sequence defined by SEQ ID NO:1 of the sequence listing together with insertion, deletion, and/or substitution by one or more amino acid residues and which have the catalytic activity of converting copalyl pyrophosphate to ent-kaurene. Particularly preferred embodiments of the polypeptide of the present invention include naturally-derived ent-kaurene synthases expressed in plant bodies such as pumpkin, maize, rice, or soybean.

25

Detailed descriptions will be given in the examples of the specification as for the evaluation of the enzymatic function of the naturally-derived ent-kaurene synthase B from pumpkin, which is a preferred embodiment of the polypeptide of the present invention. Accordingly, it can be readily understood that an ordinarily skilled artisan can easily determine whether or not a variant polypeptide, having insertion, deletion, and/or substitution of one or more constituent amino acid residues of the amino acid sequence of said enzyme, has the desired activity.

30

According to another embodiment of the present invention, polypeptides are provided which contain in their molecules amino acid sequences of the polypeptide of the naturally-derived ent-kaurene synthase B from pumpkin or variant polypeptides thereof and which have catalytic activities of converting copalyl pyrophosphate to ent-kaurene. These polypeptides are characterized by the entire sequences of the above-described polypeptides contained as a part thereof.

35

According to further embodiments of the present invention, polypeptides having catalytic activities of converting copalyl pyrophosphate to ent-kaurene are also provided which consist of, or alternatively comprise as a part thereof, a partial amino acid sequence of the polypeptide of the naturally-derived ent-kaurene synthase B from pumpkin or its variant polypeptide wherein said partial sequence has the catalytic activity of converting copalyl pyrophosphate to ent-kaurene, i.e., being a so-called active domain. Such peptides may consist solely of the active domain, per se, or may comprise the active domain as a part thereof.

40

For example, the naturally-derived ent-kaurene synthase B obtained from pumpkin, which is a preferred embodiment of the present invention, may occasionally be retained inside the cells after expression. A membrane-binding domain (a hydrophobic domain) may be selectively removed from the entire amino acid sequence of the enzyme and the remaining polypeptide containing an active domain can be ligated to a signal peptide to produce an extracellularly releasable soluble enzyme. For the determination of the hydrophobic domain, methods well known in the art, e.g., the method of Kyte et al. (Kyte, J. et al., J. Mol. Bio., 157, pp-105-132, 1982), the Hopp & Woods method and the like can be employed.

45

According to further embodiments of the present invention, DNAs which encode any one of the above-described polypeptides are provided. For example, the present invention encompasses DNAs encoding the aforementioned naturally-derived ent-kaurene synthase B from pumpkin, and as a preferred example, the DNA characterized by nucleotides of from No. 138 to 2507 (including the termination codon) of the nucleic acid sequence as defined by SEQ ID NO:3 of the sequence listing.

50

Furthermore, examples of DNAs that fall within the scope of the present invention include DNAs which encode a variant polypeptide having the catalytic activity of converting copalyl pyrophosphate to ent-kaurene wherein said polypeptide includes insertion, deletion, and/or substitution of one or more of the constituent amino acid residues of the amino acid sequence of the naturally-derived ent-kaurene synthase B from pumpkin; and DNAs which encode a

polypeptide having the catalytic activity of converting copalyl pyrophosphate to ent-kaurene wherein said DNA has a DNA sequence characterized by the nucleic acids of from 138 to 2507 (including termination codon) of the nucleic acid sequence as defined by SEQ ID NO:3 of the sequence listing in which one or more of the constituent nucleotides are inserted, deleted, and/or substituted.

5 DNAs which act as antisense DNAs against the gene containing the aforementioned sense DNAs also fall within the scope of the present invention. One example of such antisense DNAs is a DNA which acts as an antisense DNA against an ent-kaurene synthase B gene comprising a DNA encoding any one of the aforementioned polypeptides wherein said DNA sequence includes insertion, deletion, or substitution of one or more of the constituting nucleotides of the sequence that is complimentary to the DNA characterized by nucleotides of from No. 138 to 2507 (including the 10 termination codon) of the nucleic acid sequence defined as SEQ ID NO:3 of the sequence listing.

As for the DNA encoding the naturally-derived ent-kaurene synthase B from pumpkin, a preferred example of the DNAs of the present invention, method for the isolation thereof will be detailed in the examples of the specification. Therefore, DNAs encoding ent-kaurene synthase Bs derived from pumpkin and other plants can be readily obtained by those skilled in the art by following the methods described in the examples, or alternatively, by referring to the description 15 and optionally applying appropriate modifications or alterations to the described methods. Furthermore, it should be understood that DNAs containing any one of the aforementioned DNAs as a part thereof also fall within the scope of the present invention.

The DNAs of the present invention and antisense DNAs against said DNAs are useful for the regulation of the expression of ent-kaurene synthase B in plant that has important functions in the growths of plants and seeds. For 20 example, a transgenic plant introduced with the DNA of the present invention may have an increased height, and its promoted growth can be expected. A transgenic plant introduced with the antisense DNA may have a reduced height while the quantities of seeds and fruits can be maintained. Therefore, a plant bearable against a damage from a storm and flood can be obtained.

The present invention also provides a recombinant vector containing the aforementioned DNA and a transformant 25 cell containing the recombinant vector. Although the types of the vectors are not particularly limited so long as available to those skilled in the art, an appropriate vector should be chosen depending on a combination with a host cell to be transformed. Sorts of the host cell are also not particularly limited and any cells such as microbial cells (prokaryote), e.g., E. coli and animal or plant cells (eukaryote) can be used.

## 30 EXAMPLES

The present invention will be more specifically explained by referring to the examples. However, the scope of the present invention is not limited to the examples set out below.

### 35 Example 1 : Purification of the ent-kaurene synthase B (KSB)

#### (1) Materials and methods

##### (A) Plant materials and chemicals

40 Seeds of pumpkin (*Cucurbita maxima* L.) cv. "Riesenmelone gelb vernetzt" were obtained from van Waveren Pflanzenzucht, Rosdorf, Germany through Professor Jan Graebe. Plants were cultivated in a field in Saitama, Japan, and immature fruits were harvested between the middle of June and early July in 1992, when the cotyledons had reached about half their final length

45 Geranylgeraniol was a gift from Dr. T. Takigawa of Kurare Co. Ltd. (Kurashiki, Japan). Copalic acid (Nakano, T. and Djerassi C., J. Org. Chem., 26, pp.167-173, 1961) and natural resin "Brazil copal" (which contains copalic acid) were gifts from Prof. T. Nakano of Venezuela National Institute of Science (Caracas, Venezuela) and Prof. Y. Ichinohe and Dr. H. Sakamaki of Nihon University, respectively. [<sup>3</sup>H]Sodium borohydride (296 GBq mmol<sup>-1</sup>) was purchased from Amersham Int. (Buckinghamshire, UK). [1-<sup>3</sup>H]geranylgeraniol (74 GBq mmol<sup>-1</sup>) and [15-<sup>3</sup>H]copalol (74 GBq mmol<sup>-1</sup>) were prepared by reduction of geranylgeraniol and copalal with [<sup>3</sup>H]sodium borohydride (Bensen, R.J. and Zeevaart, J.A., J. Plant Growth Regul., 9, pp.237-242, 1990).

50 Tritium labeled geranylgeranyl pyrophosphate (GGPP) and copalyl pyrophosphate (CPP) were prepared through chlorination and subsequent pyrophosphorylation (Davisson et al., J. Org. Chem., 51, pp.4768-4779, 1986). The pyrophosphates were purified by reverse-phase chromatography on a Bond-Elut C18 cartridge (1 mL, Varian, Harbor City, CA) with stepwise elution using 20, 30, 50 and 70% (v/v) aqueous methanol. Optically pure uniconazole (Izumi, K., et al., Plant Cell Physiol., 26, pp.821-827, 1985) was obtained from Sumitomo Chemical Co.

## (B) Enzyme assay and protein assay

KSB activity was measured in the incubation mixtures consisted of enzyme solution in KPi buffer (50 mM, pH 8.0, 100  $\mu$  L) containing glycerol (10%, v/v), dithiothreitol (DTT) (2 mM), MgCl<sub>2</sub> (5 mM), uniconazole (20  $\mu$  M, Izumi, K., et al., Plant Cell Physiol., 26, pp.821-827, 1985) and [<sup>15</sup>-<sup>3</sup>H]CPP (1.0 kBq). [<sup>1-<sup>3</sup>H]GGPP (1.0 kBq) was used as a substrate for the measurement of KSAB activity (conversion from geranylgeranyl pyrophosphate (GGPP) to ent-kaurene).</sup>

The mixture was incubated at 30 °C for 30 min, and the enzyme reaction was terminated by the addition of acetone (200  $\mu$  L) and water (100  $\mu$  L). The mixture was extracted with n-hexane (400  $\mu$  L), and 300  $\mu$  L aliquots of the extract were concentrated in vacuo and applied to a silica gel plate. After developing with n-hexane, the radioactivity of the silica gel of the ent-kaurene region (*Rf* 0.6-1.0) was determined. One unit of kaurene synthase activity was defined as the production of 1 pmol of ent-kaurene per minute. The protein concentration was determined using BIO-RAD (Richmond, CA) microassay (Bradford, M.M., Anal. Biochem., 72, pp.248-245, 1976) using BSA as the standard.

## (C) Purification of KSB

All procedures were performed at 0 to 40°C. The following buffers were used: buffer A, 50 mM KPi buffer (pH 8.0) containing 2 mM DTT; buffer B, buffer A containing 10% glycerol; buffer C, 20 mM Tris-HCl buffer (pH 7.4) containing 10% glycerol and 2 mM DTT; buffer D, 1 mM KPi buffer (pH 6.0) containing 20% glycerol, 2 mM DTT and 5 mM MgCl<sub>2</sub>.

Crude enzyme extract (1 L) was prepared by the method of Graebe et al. (Graebe, J.E., et al., Phytochem., 13, pp.1433-1440, 1974). Butyl-Toyopearl 650S gel (Toso, Tokyo, Japan) was packed into a glass column (AP-1, 10 mm X 100 mm, Waters, Millipore Corp., Milford, MA) and the column was equilibrated with a buffer A containing 1.7 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. Ammonium sulfate was added to the crude enzyme preparation (77 mL) to 1.7 M of a final concentration, and stirred for 10 min. The enzyme solution was centrifuged at 2,000 g for 10 min, and the supernatant was loaded onto the column at a flow rate of 3.0 mL/min. The column was eluted with a 30-min linear gradient of 1.7 to 0 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and then with buffer B for 45 min. The fraction volume was 6 mL and a 5  $\mu$  L aliquot of each fraction was assayed.

The major active fractions (No. 32-37) were concentrated by ultrafiltration (YM-10 filter, Amicon, Beverly, MA) and the active fractions from several runs were pooled (53 mL). The butyl-purified enzyme preparation equilibrated with buffer C (8.4 mL) was loaded onto a DEAE-8HR column (10 mm X 190 mm, Waters). The column was eluted with a 50-min linear gradient of 0 to 500 mM NaCl in buffer C. The flow rate was 1.0 mL/min and the fraction volume was 2.0 mL.

The active fractions (No. 33-38) were pooled and concentrated by ultrafiltration (YM-10). Ammonium sulfate was added to the DEAE-purified enzyme preparation (7.9 mL) to 1.7 M of a final concentration. The enzyme solution was loaded onto a TSK phenyl-5PW column (7.5 mm X 75 mm, Toso), which was equilibrated with buffer B containing 1.7 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. The column was eluted with a 60-min linear gradient of 1.7 to 0 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and then with buffer B for 30 min. The flow rate was 1.0 mL/min and 2.0 mL fractions were collected. The major KSB activity was found in fractions No. 37-41. For the second DEAE ion exchange chromatography, the phenyl-purified enzyme preparation (2.0 mL) was charged to the DEAE column and eluted as described above.

The fractions containing most of the KSB activity (No. 36-38) were concentrated to 300  $\mu$  L by ultrafiltration (Centricon 30, Amicon). This preparation was charged to a hydroxyapatite column (TONEN TAPS-050805 HG, 8.0 mm X 50 mm, Tonen, Tokyo, Japan) equilibrated with buffer D. The column was eluted with a 60-min linear gradient of 1 to 500 mM phosphate. The flow rate was 1.0 mL/min and the fraction volume was 1 mL. The major KSB active fractions (No. 23-25) were concentrated by ultrafiltration (Centricon 30) and stored at -80°C (70  $\mu$  g, 673 units).

## (D) Gel permeation chromatography

The phenyl-purified enzyme preparation was charged onto the gel permeation column (TSK G3000SWx1, 7.8 mm x 300 mm, Toso, Tokyo, Japan) which was pre-equilibrated with buffer B containing 200 mM NaCl. The column was eluted with the same buffer at a flow rate of 0.4 mL/min and fractions were collected every 30 sec.

## (E) SDS-PAGE

Active enzyme fractions of the sequential purification steps were analyzed by SDS-PAGE using a 7.5% (w/v) gel (Laemmli, U.K., Nature, 227, pp.680-685, 1970). Approximately 3  $\mu$  g of total protein was applied to each lane. After electrophoresis, proteins were visualized by Coomassie Brilliant Blue R250 (CBB) staining. Ovalbumin (45.0 kD), BSA (66.2 kD), phosphorylase B (97.4 kD), B-galactosidase (116 kD), and myosin (200 kD) were used as standard proteins (BIO-RAD).

## (F) Properties of KSB

When the properties of KSB were examined, the phenyl-purified enzyme preparation (0.47  $\mu$  g of total protein) was

used as the enzyme solution. The amount of [ $^3\text{H}$ ]CPP was 2.2 kBq and the incubation period was 15 min. The other conditions were the same as those for the enzyme assay. To determine the optimal pH, KPi buffer (50 mM) adjusted to a particular pH (pH 4.5 to 9.0) was used. To determine the  $K_m$  value, the phenyl-purified enzyme preparation (1.2  $\mu\text{g}$  of total protein) was used, and the concentrations of [ $^3\text{H}$ ]CPP were in the range of 7 to 480 nM. The production of ent-kaurene was calculated and the  $K_m$  value was determined from a Lineweaver-Burk plot.

5

(G) N-terminal amino-acid and internal peptide sequence analysis

The final purified enzyme preparation (14.5  $\mu\text{g}$ ) was purified on a 7.5% (w/v) polyacrylamide gel. Proteins were  
10 electroblotted to a polyvinylidene difluoride membrane and visualized (Matsudaira, P., J. Biol. Chem., 262, pp.10035-10038), 1987). The 81 kD band on the membrane was cut out and the amino-acid sequence of the protein was analyzed by a protein sequencer (477A protein sequencer, Applied Biosystems Inc., Foster City, CA).

For internal peptide sequence analysis, the 81 kD protein was digested *in situ* (Kawasaki, H., et al., Anal. Biochem.,  
191, pp.332-336, 1990) with the Acromobacter protease I (a gift from Dr. T. Masaki of Ibaraki University). The generated  
15 peptide fragments were extracted from the gel and separated by reverse phase HPLC using a Superspher RP-Select B column. Aqueous 0.09% (v/v) TFA and 80% (v/v) acetonitrile containing 0.075% (v/v) TFA were used as eluents A and B, respectively. The flow rate was 0.2 mL/min and a 32-min linear gradient of 0 to 80% of eluent B was used (1090M HPLC system, Hewlett Packard, Avondale, PA). Absorbances at 215, 254, 275 and 290 nm were monitored. Amino acid sequences of separated peptides were analyzed with the protein sequencer.

20

(H) Polyclonal antibodies for the 81 kD protein

A major peptide fragment of 81 kD protein, ASQIITHPDESVLENINSW T (shown in the single character code), was  
25 custom synthesized by Quality Controlled Biochemicals Inc. (Hopkinton, MA). The synthetic peptide (3.1 mg, 1.4 nmol)  
was combined to keyhole limpet hemocyanin by using an Imjet Immunogen EDC Conjugation Kit (Pierce, Rockford, IL)  
according to the manufacturer's instructions. The combined protein (1 mg) was injected into a rabbit three times with  
intervals of 2 weeks, and 4 weeks after the last injection antiserum was collected by centrifugation of the blood. Immuno-  
30 globulin G was purified from the antiserum using a HiTrap Protein G column (1 mL, Pharmacia LKB biotechnology,  
Uppsala, Sweden).

About 10  $\mu\text{L}$  of each fraction of the 2nd DEAE-chromatography was resolved on two 7.5% SDS-PAGE gels. One  
35 of the gels was stained with CBB solution, and proteins on the other gel were electroblotted onto a nitrocellulose membrane. The nitrocellulose membrane was treated with the IgG at room temperature for 2 hr and then with alkaline phosphatase conjugated anti-rabbit-IgG at the same temperature for another 2 hr. The membrane was treated with nitro-blue tetrazolium chloride and 5-bromo-4-chloro-3'-indolylphosphate (Pierce) to visualize proteins.

35

(2) Results

(A) Purification of KSB

40 The crude enzyme prepared from immature seeds was first purified by a hydrophobic interaction chromatography in order to remove metal ions and GAs, and to concentrate the enzyme. The majority of the KSB activity was eluted at an  $(\text{NH}_4)_2\text{SO}_4$  concentration of 0.6-0.1 M, whereas the majority of the KSAB activity was eluted after the  $(\text{NH}_4)_2\text{SO}_4$  gradient was complete (Fig. 1). Fractions corresponding to an  $(\text{NH}_4)_2\text{SO}_4$  concentration of 0.6-0.1 M were collected as the butyl-purified KSB preparation, although this preparation contained some KSA activity.

45 The butyl-purified KSB preparation was purified further by DEAE ion exchange chromatography with a NaCl gradient. Both KSA and KSB co-eluted at fractions corresponding to 240-360 mM NaCl. Separation of KSA and KSB was not achieved by DEAE chromatography but the specific activity of KSB was increased (Table 1). These fractions were collected and were further purified by high performance hydrophobic interaction chromatography using a Phenyl-5PW column. The majority of KSB activity eluted at  $(\text{NH}_4)_2\text{SO}_4$  concentrations of 0.5-0.2 M, and KSAB activity eluted later.

50 About 40% of the phenyl-purified enzyme preparation was further purified on the same DEAE column. The main peak of KSB activity was detected in fractions of 250-280 mM NaCl (Fig. 2), which were collected as the second DEAE-purified enzyme preparation. This enzyme preparation contained four major proteins by SDS-PAGE, and was further purified on a hydroxyapatite column. The KSB activity was eluted at phosphate concentrations of 90-140 mM, and the peak shape of KSB activity was corresponded with that of UV-absorbance at 280 nm (Fig. 3). The purest KSB preparation did not show KSAB activity suggesting that KSA was not present in the preparation. ent-Kaurene synthase B was purified 291-fold from endosperm of *Cucurbita maxima* L. (Table 1) and 70  $\mu\text{g}$  of purified enzyme was obtained.

Table 1

Enzyme preparation	Total protein (mg)	Total activity (units)	Yield (%)	Specific activity (units/mg)	Purification (fold)
Crude enzyme	530	17,500	100	33.0	1
Butyl-purified	168	11,000	63.8	65.4	1.98
1st DEAE-purified	13.9	5,170	29.6	372	11.3
Phenyl-purified	4.43	3,340	19.2	754	22.8
2nd DEAE-purified	0.407	2,820	16.2	6,930	210
HA-purified	0.070	673	3.87	9,610	291

## (B) SDS-PAGE analysis and molecular mass estimation

The hydroxyapatite-purified enzyme preparation showed a major band at apparent molecular mass of 81 kD. The amount of protein corresponding to the 81 kD band estimated by SDS-PAGE in each fraction of the second DEAE chromatography correlated with KSB activity. To determine the molecular mass of KSB by means of gel permeation HPLC, we used a TSK G3000 SWx1 column. On this column, the elution profile of KSB was broad but centered on retention time of 23.5 min, which corresponds to an apparent molecular mass of approximately 90 kD.

## (C) Properties of KSB

The properties of KSB were examined with the phenyl-purified enzyme preparation. The optimal pH for KSB was 6.8-7.5 in 50mM KPi buffer. The  $K_m$  value for [ $^{15}$ 3H]CPP was 0.35  $\mu$ M. In the presence of 5 mM Mg<sup>2+</sup> KSB could convert CPP to ent-kaurene, but was not active without exogenous cations. Since EDTA (1-100 mM) did not change KSB activity, the phenyl-purified enzyme preparation did not contain endogenous cations that affected KSB activity.

Some divalent cations (0.1-50mM, as chlorides) were added to the incubation mixture instead of 5 mM Mg<sup>2+</sup>, and KSB activity was measured. As shown in Fig. 4, Mg<sup>2+</sup>, Co<sup>2+</sup> and Mn<sup>2+</sup> promoted KSB activity and their optimal concentrations were around 10 mM. Ni<sup>2+</sup> and Fe<sup>2+</sup> weakly promoted KSB activity, while Cu<sup>2+</sup>, Ca<sup>2+</sup> and Ba<sup>2+</sup> showed no effects. When Cu<sup>2+</sup>, Ca<sup>2+</sup> and Ba<sup>2+</sup> were added to the incubation mixture in combination with 5 mM Mg<sup>2+</sup>, all three inhibited KSB activity promoted by Mg<sup>2+</sup>. In particular Cu<sup>2+</sup> completely inhibited the conversion promoted by 5 mM Mg<sup>2+</sup> at concentrations above 0.5 mM.

## (D) Peptide sequence analysis

The N-terminal peptide sequence of the 81 kD protein from SDS-PAGE was analyzed and the N-terminus of the protein was found to be blocked. The 81 kD protein band was cut out from the polyacrylamide gel and treated with a lysyl endopeptidase, Acromobacter protease I (Masaki, T., et al., Biochem. Biophys. Acta, 660, pp.44-50, 1981). The resulting peptides were separated by reverse-phase HPLC and the sequences of the major peptides analyzed (Kawasaki, H., et al., Anal. Biochem., 191, pp.332-336, 1990). Sequences and approximate amounts of the peptides recovered are (1) ASQIITHPDESVLENINSWT (74 pmol), (2) EAEWSTNK (45 pmol), (3) RAMESYSGDIVRISK (37 pmol), (4) HGLSSDSVW (24 pmol), (5) LQDWMSVMQYQRK (17 pmol).

## (E) Western blot analysis and immunoprecipitation

Western blot analysis showed that the polyclonal antibody raised against the synthetic peptide ASQIITHPDESVLENINSWT, combined to keyhole limpet hemocyanin, selectively recognized the 81 kD protein in each fraction from the 2nd DEAE-chromatography, and the pattern showed a good correlation with the patterns of KSB activity and CBB staining. In an immunoprecipitation experiment using the phenyl-purified enzyme and the polyclonal antibody, the 81 kD protein was detected in the supernatant fraction and not in the precipitate regardless of the amount of antibody used.

## Example 2: Cloning of ent-kaurene synthase gene

## (A) Isolation of the KSB cDNA clone

5 From the peptide sequences ([1] ASQIITHPD ESVLENINSWT, (2) EAEWSTNK, [3] RAMESYSGDIVRISK, [4] HGLSSDSVW, [5] LQDWSMVMQYQRK), degenerated oligonucleotides were synthesized for PCR primers in both sense and antisense directions. Using these synthetic oligonucleotides as primers, (RT)-PCR was tried to amplify cDNA fragments from poly(A)<sup>+</sup> RNA prepared from immature cotyledons of *Cucurbita maxima*.

10 As the result, a 472-bp cDNA fragment was amplified when a combination of the forward primer designed from the underlined part of the sequence [5] and the reverse primer from the underlined sequence in [1] was used as a primer set. The amino acid sequence showed relatively high homology with KSA of *Arabidopsis* (39% identity, 66% similarity). Because the amino acid sequence of this region of putative maize KSA (Bensen, R.J. et al., Plant Cell, 7, pp.75-84, 1995) has much higher homology with *Arabidopsis* KSA, it was suggested that the PCR amplified fragment was not a part of KSA, but that of one of other terpene cyclases.

15 Using the PCR-amplified fragment as a probe, RNA blot analysis was performed against poly(A)<sup>+</sup>RNA prepared from the same material to detect a 2.7-kb transcript, a possible size compared to the estimated size of the purified protein (81 kD). Considering these results together with the fact that the final purified enzyme fraction (the hydroxyapatite purified fraction as described above) used for sequencing did not contain KSA activity but had KSB activity, it is strongly suggested that the 472-bp PCR-amplified fragment encodes a part of KSB cDNA.

20 To obtain a full-length KSB cDNA clone, a cDNA library consisting of 1.2 million independent clones was constructed from poly(A)<sup>+</sup> RNA prepared from immature cotyledons of *Cucurbita maxima* using λ ZAP II as a vector. After the amplification of the library,  $5 \times 10^5$  clones were screened to isolate 9 independent clones after 3 rounds of purifications. After the subcloning into pBluescriptSK(-), the size of the insert cDNA was determined to select 2 clones, pKB3 and pKB16 for further studies. Both of the clones contained 2.4-kb insert and showed same restriction enzyme maps.

25 Comparison of the size of the cDNA insert and that of the transcript detected in Northern blot (2.7 kb) implied that the inserts of pKB3 and pKB16 are 5'-truncated DNAs. The method of rapid amplification of cDNA ends was used for the isolation of the 5'-region of pKB16, and a 0.6-kb cDNA fragment containing an additional 263 bp in 5'-region was isolated, which was fused to pKB16 by use of PstI site locating at nucleotide 722. The resulting plasmid clone containing the fused cDNA was named pKB20.

## 30 (B) Sequencing analysis of pKB20

pKB 20 contains a cDNA insert consisting of 2658 nucleotides, which contains the identical sequence (801 to 1269) to the PCR amplified fragment which was used as a probe. The nucleotide sequence of the full length cDNA is given in SEQ ID NO:2 of the sequence listing. The KSB code region of the cDNA and the amino acid sequence of the corresponding KSB are given in SEQ ID NO:3 of the sequence listing. The amino acid sequence of KSB is shown in SEQ ID NO:1 of the sequence listing according to the IUPAC single character coding. The ATG codon at the position of 138 to 140 is most likely the starting codon for KSB protein, which is followed by a long ORF of 2367 nucleotides and polyadenylation signal sequence (AATAAA) at the position of 103 to 109 nucleotides downstream from the translational stop codon.

40 The ORF consists of 789 amino acids, where all the peptide sequences obtained from the protease-digested purified protein are detected, although there was a change of amino acid in [5] LQDWSMVMQYQRK to LQDWDM-VMQYQRK. The first 50 N-terminal amino acid sequence is rich in serine (18%) and threonine (12%) and the estimated pI of this region is 9.8, which are common characters of transit peptides targeting chloroplasts (Keegstra, et al., Annu. Rev. Plant Physiol. Mol. Biol., 40, 471-501, 1989).

45 As shown in above example 1, the polyclonal antibody raised against the synthetic oligopeptide, [1] ASQIITHP-DESVLENINSWT, combined to a carrier protein did not recognize the native KSB protein, although it recognized a 81 kD protein on SDS-PAGE. From the results, it was predicted that the peptide sequence located not on the surface, but in the core part of the properly folded protein, so that the antibody could recognize only the denatured protein. In the sequence of above, the peptide sequence used for the production of the antibody is found in just the middle of the sequence (nucleotide 1244 to 1304), not in the N-terminus nor C-terminus region, which are supporting the experimental results of Example 1.

## (c) Expression of KB20 protein in Escherichia coli

55 To confirm whether pKB20 encodes KSB, the protein encoded by the cloned cDNA was over-expressed as a fusion protein with maltose binding protein (Guan, et al., Gene, 67, 21-30, 1987) in *Escherichia coli* to analyze the enzyme activity of the expressed protein. Plasmid pMAL-KB containing a fragment (tac promoter sequence-malE-KSB) for the expression of the fusion protein (MBP-KB) and the control plasmid pMAL (42 kD for MBD expression) containing tac-

promoter sequence-malE fragment were prepared.

In this system, the expression of the fusion protein is under the regulation of tac promoter, which is induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG). On an SOS-PAGE followed by Coomassie blue staining, a band of an expected size (131 kD) of the fusion protein (MBP-KB) was detected in total protein extract of *E. coli* which carries pMALKB after IPTG induction. Immunoblot analysis was performed to confirm whether the IPTG-induced protein is recognized by the anybody against the synthetic peptide. Immuno-reactivity was detected on the band corresponding to the IPTG-induced protein, and additional bands of smaller size were also detected, which seemed to be the degradation products of the KB20 protein. However, no immuno-staining was observed in the protein of *E. coli* producing MBP only, regardless of the addition of IPTG.

10

#### (D) Functional analysis of KB20 protein

To analyze the enzyme activity of MBP-KB, soluble proteins were prepared from *E. coli* producing the fusion protein. But only a trace amount of the fusion protein was found in the soluble fraction, suggesting that the most of the fusion protein formed an inclusion body, when the *E. coli* was cultured under normal conditions. Bishai et al. (Bishai, et al., J. Bact. 169, pp.5140-5151, 1987) reported that when *E. coli* is cultured at a low temperature after the induction by IPTG, the ratio of the over-expressed fusion protein in a fraction of soluble proteins increased. In the case of MBP-KB, at 20°C, the ratio of MBP-KB protein in soluble fraction raised to about 10 folds compared to *E. coli* cultured at 37 °C. The temperature fits a physiological condition of the plant which is the origin of the enzyme, thus is seemed to help the stability of the enzyme to culture *E. coli* at a low temperature (20°C).

The conversion to ent-kaurene was analyzed using the soluble protein extract of *E. coli* grown at 20 °C after the addition of IPTG, according to the KSB enzyme assay as the above Example 1 and adding [<sup>3</sup>H]CPP to the extracts of *E. coli* which carries pMALKB or pMAL. After the incubation of [<sup>3</sup>H]CPP with the *E. coli* extracts in the presence of Mg<sup>2+</sup> (cofactor for KSB of pumpkin) and uniconazole (an inhibitor of ent-kaurene oxidation), the reaction mixture was extracted with n-hexane followed by fractionation on silica-gel thin-layer chromatography (TLC) to isolate LP (low polarity) fraction, where ent-kaurene is fractionated.

KSB activity was tentatively defined as the radioactivity (dpm) of the "LP fraction". The extract of *E. coli* producing MBP-KB20 showed significant KSB activities, although that of the bacteria cells expressing only MBP had no activities. It was analyzed on OSD-HPLC to confirm whether the radioactivity in "LP fraction" on TLC corresponds to ent-kaurene. The retention time of the major peak in LP fraction of the *E. coli* extract was identical to that of the original pumpkin enzyme preparation, besides, those of them are identical to retention time of authentic [<sup>14</sup>C]ent-kaurene. It was also proved that the radioactivity in LP fraction of the *E. coli* extracts were mostly derived from the major peak (95%) identical to ent-kaurene fraction. Thus, it was indicated that the extract of *E. coli* producing MBP-KB had enzyme activity to convert CPP to ent-kaurene.

According to the results of Northern blotting, KSB transcripts were detected in all of the tested tissues and this fact indicates that the KSB could be expressed not only in immature seeds but in vegetative tissues. On the other hand, a message of 20-oxidase of pumpkin is detected only in immature seeds in contrast to KSB (Lange, T., et al., Proc. Natl. Acad. Sci. USA, 91, pp.8552-8556, 1994). The expression level of KSB is relatively high in young tissues and the transcripts were detected most abundantly in growing cotyledon on growth of immature seeds.

40

#### (E) Sequence comparison with other terpene cyclases

The deduced amino acid sequence of KB20 and sequences of other terpene cyclases, including KSAs, was analyzed with comparison to each other. KB20 sequence shared homology with tobacco sesquiterpene cyclase (Facchini, P.J., Proc. Natl. Acad. Sci. USA, 89, pp.11088-11092, 1992), limonene synthase (Colby, S.M., J. Biol. Chem., 268, pp.23016-23024, 1993), castor bean casbene synthase (Colby, S.M., J. Biol. Chem., 268, pp.23016-23024, 1993) and in particular, *Arabidopsis* KSA (Sun T.P. and Kamiya, Y., Plant Cell, 6, pp.1509-1518, 1994).

Around the center part of the sequences (amino acid 337 to 600) are well conserved in all the terpene cyclases analyzed (Sun T.P. and Kamiya, Y., Plant Cell, 6, pp.1509-1518, 1994). *Arabidopsis* KSA and KSB20 shared the identical amino acids in the first part of the sequences (amino acid 114 to 230, 34.8% identity), although common amino acid between the KSA and the KSB are hardly observed in the last part of the sequences (amino acid 601 to the terminus). However, KSB and cyclases other than KSA shared some conserved amino acid in the region of amino acid 603 to 631.

55

SEQUENCE LISTING

5

(1) GENERAL INFORMATION:

10

- (i) APPLICANT: INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH  
(ii) TITLE OF INVENTION: KAURENE SYNTHASE  
(iii) NUMBER OF SEQUENCES: 3

15

(2) INFORMATION FOR SEQ ID NO:1:

20

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 789 amino acids  
(B) TYPE: amino acid  
(C) TOPOLOGY: linear  
(ii) MOLECULE TYPE: protein

25

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:1:

30

MYLSRPTGVARFAASSSSSSASLFFPGVDVDTTGTALHFEETKERIKKLFDKVELSVS  
AYDTAWVAMVPSPNSLNQPLFPECINWVLDSQHADGSWGLLHNDQLLMKANLLSTLACVL  
TLKRWNIGHDEMSKALDFIKSNIASATDENQRSPVGFDIIPPGMIEYAKDLNLNLPLAPT  
NVDALVRKKELELRSCRSNSEGGKAYLAYVSEGIGKLQDWDVMVQYQRKNGLFNSPSTT  
AAAFMHRNDDGCPDYLRSLIQKFDSVPTIYPLDIYARLHMVDSLQKFGIARHFKEEIRS  
VLDETYRCWMQGEENIFLDASTCAMAFLRVEGYDVSSDQLTQFSEDIPPNCLGGYLKD  
40  
FGASLELYKASQIIITHPDESVLENINSWTSRPLKHGLSSDSVWSDRDSVVKQEAVNALE  
FPYNATLERLISKAMESYSGDIVRISKSPYACLNFGHQDFLELAVEDFNTLQRIHLKEL  
EELQRWVVENKLDDELKPFRHLGYCYFAAAATLTDPELHDARIWAQNGVLTTVVDDFYD  
45  
GGGSEEELDNLIELVEKWDPDGEGVGYCSKDVEIVFLALHSTVCEIGRRALWQGRSVMRN  
VIDGWLALLKVMRKEAEWSTNKVVPSMGEYMEQAHSFALGPIILPMLPPVGPKLSSEMI  
50  
GSCEYQKLYKLMSTAGRLKNDIRSYDRECKEGKLNILSLWMIDGGGNVTKEEAIEAIKGD

55

FERAIRELLGLVLQENTTIPRACKDLFWKLMIVNLFYMEDDGYSNRLMNTVKAMFEQP

MDLDALLNK 789

5

## (2) INFORMATION FOR SEQ ID NO:2:

10

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2658 base pairs

15

(B) TYPE: nucleic acid

(C) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: cDNA

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:2:

25

GAAAGCGGTA GTTTTCGGCT CTTGGCTGTT TTTGTGTACT GCCTATTCTG TTCTTTATAT  
 GCACTGAATG TGAAAATGCC TCACCCAGAGT TTGATGCTGC TTCCGCAATT TCTTTGCATG  
 TCCTAACAT ACCAGTCATG TATCTTCCC GACCTACCGG CGTTGCCCGT TTTGCTGCCT  
 CCTCCTCCTC CTCCTCCTCA GCTTCCCTAT TCCCTGGAGT GGATGTGGAC ACAACAACCA  
 AAACCTGGAGC TTTGCACTTT GAAGAAACAA AAGAAAGAAT TAAGAAATTG TTGACAAGG  
 TTGAACCTTC AGTTCTGCA TATGATACTG CATGGGTGGC AATGGTTCCCT TCTCCAAACT  
 CTCTCAACCA ACCTCTTTTC CCCGAGTGTA TAAACTGGGT ATTAGATAGT CAACATGCTG  
 ATGGCTCATG GGGCCTACTC CACAACGATC AGTTGCTGAT GAAGGCCAAT CTCTTATCTA  
 CATTAGCATG TGTTCTTACT CTTAACCGGTT GGAATATTGG GCATGATCAT ATGAGCAAGG  
 CCCTTGATTT TATCAAGTCT AATATAGCTT CAGCTACTGA TGAGAACCAA CGTTCTCCGG  
 TGGGATTTGA CATTATTTTC CCTGGCATGA TTGAGTATGC TAAAGACTTG AATTTGAATC  
 TACCCCTGGC ACCGACGAAC GTGGATGCCT TGGTTGAAA GAAAGAGTTG GAGCTGAGAA  
 GCTGCAGAAG CAACTCTGAA GGTGGAAAAG CCTATTTAGC GTATGTTCA GAAGGAATTG  
 50 GAAAGTTACA GGACTGGGAT ATGGTCATGC AATATCAAAG GAAGAATGGA TCACTGTTA  
 ATTCTCCATC CACTACGGCA GCGGCTTTA TGATGATGGC TGTTTGATT

55

ATCTTCGCTC ACTCTTACAA AAGTTTGATG GCTCAGTTCC CACAATATAT CCTCTTGATA  
 5 TATATGCTCG ATTACACATG GTTGATAGCC TTCAAAATT CGGAATTGCT CGGCATTCA  
 AAGAGGGAGAT TAGAAGCGTA TTAGATGAAA CTTACAGGTG TTGGATGCAA GGAGAGGAAA  
 10 ATATATTCTT AGATGCTTCA ACTTGTGCAA TGGCTTTG AATGTTACGT GTTGAAGGAT  
 ATGATGTTTC TTCAGACCAAG TTGACTCAAT TTTAGAAGA TATCTTCCC AATTGCCTTG  
 15 GAGGATATTT AAAAGACTTC GGTGCCTCGC TGGAGTTATA TAAGGCCTCT CAGATTATCA  
 CGCACCCCCGA TGAATCTGTT CTGGAAAATA TAAACTCTTG GACTAGTCGT TTCTGAAGC  
 ATGGATTATC TAGTGATTCA GTTGGTCTG ATAGAACCGA TAGTGTGTT AAACAAGAGG  
 20 CTGTTAATGC TCTTGAGTTTC CCCTATAATG CAACTCTAGA ACGCCTAATA AGTAAGAGGG  
 CAATGGAAAG TTACAGTGGA GACATTGTGA GGATTCAAA ATGCCATAT GCCTGCTTAA  
 25 ATTTTGGCCA TCAAGATTTT CTGGAACCTG CTGTAGAGGA TTTCAATACC CTGCAACGCA  
 TTCATCTTAA AGAACCTGGAA GAGCTTCAAA GATGGGTGGT TGAAAACAAA TTGGACGAGT  
 TGAAATTTT CAGACTGCAC CTAGGGTACT GCTATTTGC TGCGGCAGCG ACCCTTACTG  
 30 ATCCCTGAAC TCATGATGCT CGCATAGCAT GGGCACAAAA TGGTGTGCTC ACGACCGTGG  
 TTGATGATT CTATGATGGT GGAGGATCTG AAGAGGAATT GGATAACCTT ATAGAATTGG  
 TGGAAAAGTG GGATCCTGAT GGGGAAGTGG GTTACTGTTT CAAGGACGTT GAGATTGTAT  
 35 TTCTTGCACT GCACAGCACA GTTGTGAAA TAGGAAGAAG AGCTTAGTA TGGCAAGGAC  
 GCAGTGTAT GAGGAATGTT ATCGATGGTT GGTTGGCTCT GCTGAAGGTG ATGAGAAAGG  
 AAGCTGAATG GTCGACAAAT AAGGTAGTGC CATCAATGGG TGAATATATG GAACAAGCCC  
 40 ATGTATCATT CGCGTTGGGA CCTATAATCC TTCCAATGCT CTTCTTGTT GGACCTAAC  
 TCTCAGAGGA AATGATTGGA AGCTGTGAAT ACCAGAAGTT ATATAAGCTG ATGAGCACTG  
 45 CTGGTCGCCT TAAGAATGAT ATTGATCTT ACGATAGAGA ATGCAAAGAG GGAAAGCTGA  
 ATATTCTGTC TCTGTGGATG ATTGATGGCG GTGGTAATGT CACCAAAGAG GAGGCCATTG  
 AAGCAATTAA AGGGGATTTT GAGAGGGCGA TAAGAGAGCT GCTGGGGTTA GTTTGCAGG  
 50 AGAACACTAC AATTCCAAGA GCTTGTAAAGG ATTTGTTCTG GAAATTGATG TCCATTGTGA  
 ATCTATTTA CATGGAAGAT GATGGGTACA CTTCAAATAG GTTGATGAAC ACTGTAAAAG

5 CCATGTTGA ACAACCCATG GATCTGGATG CACTATTGAA CAAATGAACG AACAAACAAAC  
 10 GAAATAAGTT GTAGTGGAA GCAGACGCCA CCTATTCACT GTCTTCTCTA CTGATCAGAT  
 15 GTCTATTGCT TTCTCATTGA GGAAATGGCA AATAAAACAT TTTATGGAAA CATTTCATTG  
 20 TTAAATTGAG ACAGTCTG

10

## (2) INFORMATION FOR SEQ ID NO:3:

15

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2658 base pairs
- (B) TYPE: nucleic acid
- 20 (C) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

25

## (iii) SEQUENCE DESCRIPTION:SEQ ID NO:3:

GAAAGCGGTA GTTTTCGGCT CTTGGCTGTT TTTGTGTACT GCCTATTCTG TTCTTTATAT	60
GCACTGAATG TGAAAATGCC TCACCCAGAGT TTGATGCTGC TTCCGCAATT TCTTTGCATG	120
30 TCCTAAACAT ACCAGTC	137

35

ATG TAT CTT TCC CGA CCT ACC GGC GTT GCC CGT TTT GCT GCC TCC	182
Met Tyr Leu Ser Arg Pro Thr Gly Val Ala Arg Phe Ala Ala Ser	

40

TCC TCC TCC TCC TCC TCA GCT TCC CTA TTC CCT GGA GTG GAT GTG	227
Ser Ser Ser Ser Ser Ala Ser Leu Phe Pro Gly Val Asp Val	

45

GAC ACA ACA ACT AAA ACT GGA GCT TTG CAC TTT GAA GAA ACA AAA	
Asp Thr Thr Thr Lys Thr Gly Ala Leu His Phe Glu Glu Thr Lys	
50	

55

GAA AGA ATT AAG AAA TTG TTC GAC AAG GTT GAA CTT TCA GTT TCT	317
---	-----

EP 0 768 381 A2

Glu Arg Ile Lys Lys Leu Phe Asp Lys Val Glu Leu Ser Val Ser

5 GCA TAT GAT ACT GCA TGG GTG GCA ATG GTT CCT TCT CCA AAC TCT  
Ala Thy Asp Thr Ala Trp Val Ala Met Val Pro Ser Pro Asn Ser

10 CTC AAC CAA CCT CTT TTC CCC GAG TGT ATA AAC TGG GTA TTA GAT 407  
Leu Asn Gln Pro Leu Phe Pro Glu Cys Ile Asn Trp Val Leu Asp

20 AGT CAA CAT GCT GGC TCA TGG GGC CTA CTC CAC AAC GAT CAG  
Ser Gln His Ala Asp Gly Ser Trp Gly Leu Leu His Asn Asp Gln

25 TTG CTG ATG AAG GCC AAT CTC TTA TCT ACA TTA GCA TGT GTT CTT 497  
Leu Leu Met Lys Ala Asn Leu Leu Ser Thr Leu Ala Cys Val Leu

30 ACT CTT AAG CGG TGG ATT ATT GGG CAT GAT CAT ATG AGC AAG GCC  
Thr Leu Lys Arg Trp Asn Ile Gly His Asp His Met Ser Lys Ala

35 CTT GAT TTT ATC AAG TCT AAT ATA GCT TCA GCT ACT GAT GAG AAC 587  
Leu Asp Phe Ile Lys Ser Asn Ile Ala Ser Ala Thr Asp Glu Asn

45 CAA CGT TCT CCG GTG GGA TTT GAC ATT ATT TTC CCT GGC ATG ATT  
Gln Arg Ser Pro Val Gly Phe Asp Ile Ile Phe Pro Gly Met Ile

50 GAG TAT GCT AAA GAC TTG AAT TTG AAT CTA CCC TTG GCA CCG ACC 677  
Glu Tyr Ala Lys Asp Leu Asn Leu Asn Leu Pro Leu Ala Pro Thr

EP 0 768 381 A2

AAC GTG GAT GCC TTG GTT CGA AAG AAA GAG TTG GAG CTG AGA AGC  
5 Asn Val Asp Ala Leu Val Arg Lys Lys Glu Leu Glu Leu Arg Ser

TGC AGA AGC AAC TCT GAA GGT GGA AAA GCC TAT TTA GCG TAT GTT 767  
10 Cys Arg Ser Asn Ser Glu Gly Gly Lys Ala Tyr Leu Ala Tyr Val

TCA GAA GGA ATT GGA AAG TTA CAG GAC TGG GAT ATG GTC ATG CAA  
15 Ser Glu Gly Ile Gly Lys Leu Gln Asp Trp Asp Met Val Met Gln

TAT CAA AGG AAG AAT GGA TCA CTG TTT AAT TCT CCA TCC ACT ACG 857  
20 Tyr Gln Arg Lys Asn Gly Ser Leu Phe Asn Ser Pro Ser Thr Thr

GCA GCG GCT TTT ATG CAT AGA AAT GAT GAT GGC TGT TTT GAT TAT  
25 Ala Ala Ala Phe Met His Arg Asn Asp Asp Gly Cys Phe Asp Tyr

30 CTT CGC TCA CTC TTA CAA AAG TTT GAT GGC TCA GTT CCC ACA ATA 947  
Leu Arg Ser Leu Leu Gln Lys Phe Asp Gly Ser Val Pro Thr Ile

TAT CCT CTT GAT ATA TAT GCT CGA TTA CAC ATG GTT GAT AGC CTT  
40 Tyr Pro Leu Asp Ile Tyr Ala Arg Leu His Met Val Asp Ser Leu

CAA AAA TTC GGA ATT GCT CGG CAT TTC AAA GAG GAG ATT AGA AGC 1037  
45 Gln Lys Phe Gly Ile Ala Arg His Phe Lys Glu Glu Ile Arg Ser

GTA TTA GAT GAA ACT TAC AGG TGT TGG ATG CAA GGA GAG GAA AAT  
50 Val Leu Asp Glu Thr Tyr Arg Cys Trp Met Gln Gly Glu Glu Asn

55

## EP 0 768 381 A2

ATA TTC TTA GAT GCT TCA ACT TGT GCA ATG GCC TTT CGA ATG TTA	1127
5 Ile Phe Leu Asp Ala Ser Thr Cys Ala Met Ala Phe Arg Met Leu	
CGT GTT GAA GGA TAT GAT GTT TCT TCA GAC CAG TTG ACT CAA TTT	
10 Arg Val Glu Gly Tyr Asp Val Ser Ser Asp Gln Leu Thr Gln Phe	
TCA GAA GAT ATC TTT CCC AAT TGC CTT GGA GGA TAT TTA AAA GAC	1217
15 Ser Glu Asp Ile Phe Pro Asn Cys Leu Gly Gly Tyr Leu Lys Asp	
TTC GGT GCC TCG CTG GAG TTA TAT AAG GCC TCT CAG ATT ATC ACG	
20 Phe Gly Ala Ser Leu Glu Leu Tyr Lys Ala Ser Gln Ile Ile Thr	
CAC CCC GAT GAA TCT GTT CTG GAA AAT ATA AAC TCT TGG ACT AGT	1307
25 His Pro Asp Glu Ser Val Leu Glu Asn Ile Asn Ser Trp Thr Ser	
CGT TTC CTG AAG CAT GGA TTA TCT AGT GAT TCA GTT TGG TCT GAT	
30 Arg Phe Leu Lys His Gly Leu Ser Ser Asp Ser Val Trp Ser Asp	
AGA ACC GAT AGT GTT GTT AAA CAA GAG GCT GTT AAT GCT CTT GAG	1397
40 Arg Thr Asp Ser Val Val Lys Gln Glu Ala Val Asn Ala Leu Glu	
TTC CCC TAT AAT GCA ACT CTA GAA CGC CTA ATA AGT AAG AGG GCA	
45 Phe Pro Tyr Asn Ala Thr Leu Glu Arg Leu Ile Ser Lys Arg Ala	
50 ATG GAA AGT TAC AGT GGA GAC ATT GTG AGG ATT TCA AAA TCG CCA	1487

55

EP 0 768 381 A2

Met Glu Ser Tyr Ser Gly Asp Ile Val Arg Ile Ser Lys Ser Pro

5

TAT GCC TGC TTA AAT TTT GGC CAT CAA GAT TTT CTG GAA CTT GCT  
Tyr Ala Cys Leu Asn Phe Gly His Gln Asp Phe Leu Glu Leu Ala

10

GTA GAG GAT TTC AAT ACC CTG CAA CGC ATT CAT CTT AAA GAA CTG  
Val Glu Asp Phe Asn Thr Leu Gln Arg Ile His Leu Lys Glu Leu

1577

20

GAA GAG CTT CAA AGA TGG GTG GTT GAA AAC AAA TTG GAC GAG TTG  
Glu Glu Leu Gln Arg Trp Val Val Glu Asn Lys Leu Asp Glu Leu

25

AAA TTT TTC AGA CTG CAC CTA GGG TAC TGC TAT TTT GCT GCG GCA  
Lys Phe Phe Arg Leu His Leu Gly Tyr Cys Tyr Phe Ala Ala Ala

1667

30

GCG ACC CTT ACT GAT CCT GAA CTT CAT GAT GCT CGC ATA GCA TGG  
Ala Thr Leu Thr Asp Pro Glu Leu His Asp Ala Arg Ile Ala Trp

35

GCA CAA AAT GGT GTG CTC ACG ACC GTG GTT GAT GAT TTC TAT GAT  
Ala Gln Asn Gly Val Leu Thr Thr Val Val Asp Asp Phe Tyr Asp

1757

45

GGT GGA GGA TCT GAA GAG GAA TTG GAT AAC CTT ATA GAA TTG GTG  
Gly Gly Gly Ser Glu Glu Glu Leu Asp Asn Leu Ile Glu Leu Val

50

GAA AAG TGG GAT CCT GAT GGG GAA GTG GGT TAC TGT TCC AAG GAC  
Glu Lys Trp Asp Pro Asp Gly Glu Val Gly Tyr Cys Ser Lys Asp

1847

55

EP 0 768 381 A2

GTT GAG ATT GTA TTT CTT GCA CTG CAC AGC ACA GTT TGT GAA ATA  
Val Glu Ile Val Phe Leu Ala Leu His Ser Thr Val Cys Glu Ile

5

GGA AGA AGA GCT TTA GTA TGG CAA GGA CGC AGT GTT ATG AGG AAT                    1937  
10 Gly Arg Arg Ala Leu Val Trp Gln Gly Arg Ser Val Met Arg Asn

10

GTT ATC GAT GGT TGG TTG GCT CTG CTG AAG GTG ATG AGA AAG GAA  
15 Val Ile Asp Gly Trp Leu Ala Leu Lys Val Met Arg Lys Glu

20

GCT GAA TGG TCG ACA AAT AAG GTA GTG CCA TCA ATG GGT GAA TAT                    2027  
Ala Glu Trp Ser Thr Asn Lys Val Val Pro Ser Met Gly Glu Tyr

25

ATG GAA CAA GCC CAT GTA TCA TTC GCG TTG GGA CCT ATA ATC CTT  
Met Glu Gln Ala His Val Ser Phe Ala Leu Gly Pro Ile Ile Leu

30

CCA ATG CTC TTC TTT GTT GGA CCT AAA CTC TCA GAG GAA ATG ATT                    2117  
Pro Met Leu Phe Phe Val Gly Pro Lys Leu Ser Glu Glu Met Ile

35

GGA AGC TGT GAA TAC CAG AAG TTA TAT AAG CTG ATG AGC ACT GCT  
40 Gly Ser Cys Glu Tyr Gln Lys Leu Tyr Lys Leu Met Ser Thr Ala

45

GGT CGC CTT AAG AAT GAT ATT CGA TCT TAC GAT AGA GAA TGC AAA                    2207  
Gly Arg Leu Lys Asn Asp Ile Arg Ser Tyr Asp Arg Glu Cys Lys

50

GAG GGA AAG CTG AAT ATT CTG TCT CTG TGG ATG ATT GAT GGC GGT  
Glu Gly Lys Leu Asn Ile Leu Ser Leu Trp Met Ile Asp Gly Gly

55

GGT AAT GTC ACC AAA GAG GAG GCC ATT GAA GCA ATT AAA GGG GAT 2297  
 5 Gly Asn Val Thr Lys Glu Glu Ala Ile Glu Ala Ile Lys Gly Asp

TTT GAG AGG GCG ATA AGA GAG CTG CTG GGG TTA GTT TTG CAG GAG 10  
 Phe Glu Arg Ala Ile Arg Glu Leu Leu Gly Leu Val Leu Gln Glu

AAC ACT ACA ATT CCA AGA GCT TGT AAG GAT TTG TTC TGG AAA TTG 15 2387  
 20 Asn Thr Thr Ile Pro Arg Ala Cys Lys Asp Leu Phe Trp Lys Leu

ATG TCC ATT GTG AAT CTA TTT TAC ATG GAA GAT GAT GGG TAC ACT 25  
 Met Ser Ile Val Asn Leu Phe Tyr Met Glu Asp Asp Gly Tyr Thr

TCA AAT AGG TTG ATG AAC ACT GTA AAA GCC ATG TTT GAA CAA CCC 2477  
 30 Ser Asn Arg Leu Met Asn Thr Val Lys Ala Met Phe Glu Gln Pro

ATG GAT CTG GAT GCA CTA TTG AAC AAA TGA 2507  
 35 Met Asp Leu Asp Ala Leu Leu Asn Lys \*\*\*

ACGAACAACA AACGAAATAA GTTGTAGTGG GAAGCAGACCG CCACCTATTG ACTGTCTTCT 2567  
 40 CTACTGATCA GATGTCTATT GCTTTCTCAT TGAGGAAATG GCAAATAAAA CATTATGG 2627  
 45 AACACATTCA TTGTTAAATT GAGACAGTCT G 2658

## 50 Claims

1. A polypeptide which is characterized by amino acid Nos. 1-789 of the amino acid sequence as defined by SEQ ID NO:1 of the sequence listing.
- 55 2. A polypeptide which has a catalytic activity of converting copalyl pyrophosphate to ent-kaurene wherein said polypeptide has the amino acid sequence as defined by amino acid Nos. 1-789 of SEQ ID NO:1 of the sequence listing in which one or more constituent amino acid residues are inserted, deleted, and/or substituted.
3. The polypeptide according to claim 2 which is a naturally-derived enzyme expressed in a plant body.

**EP 0 768 381 A2**

4. A polypeptide which contains a polypeptide according to any one of claims 1 to 3 in the molecule thereof and has a catalytic activity of converting copalyl pyrophosphate to ent-kaurene.
5. A polypeptide which consists of or comprises as a part thereof a partial polypeptide being derived from any one of the polypeptides according to any one of claims 1 to 3 and being an active domain catalyzing the conversion of copalyl pyrophosphate to ent-kaurene.
6. A DNA which encodes the polypeptide according to claim 1.
10. 7. The DNA according to claim 6 which is characterized by the nucleotide Nos. 138-2507 of the nucleic acid sequence as defined by SEQ ID NO:3 of the sequence listing.
15. 8. A DNA which encodes a polypeptide having an enzymatic activity of converting copalyl pyrophosphate to ent-kaurene and which has the DNA sequence according to claim 7 and contains insertion, deletion, and/or substitution of one or more of the constituent nucleotides.
20. 9. An anti-sense DNA against an ent-kaurene B synthase gene comprising the DNA according to claim 6 wherein said anti-sense DNA has a DNA sequence complimentary to the DNA according to claim 7 and contains insertion, deletion, and/or substitution of one or more of constituent nucleotides.
10. A recombinant vector which contains a DNA according to any one of claims 6 to 9.
11. A transformant which contains a recombinant vector according to claim 10.

25

30

35

40

45

50

55

Fig.1.

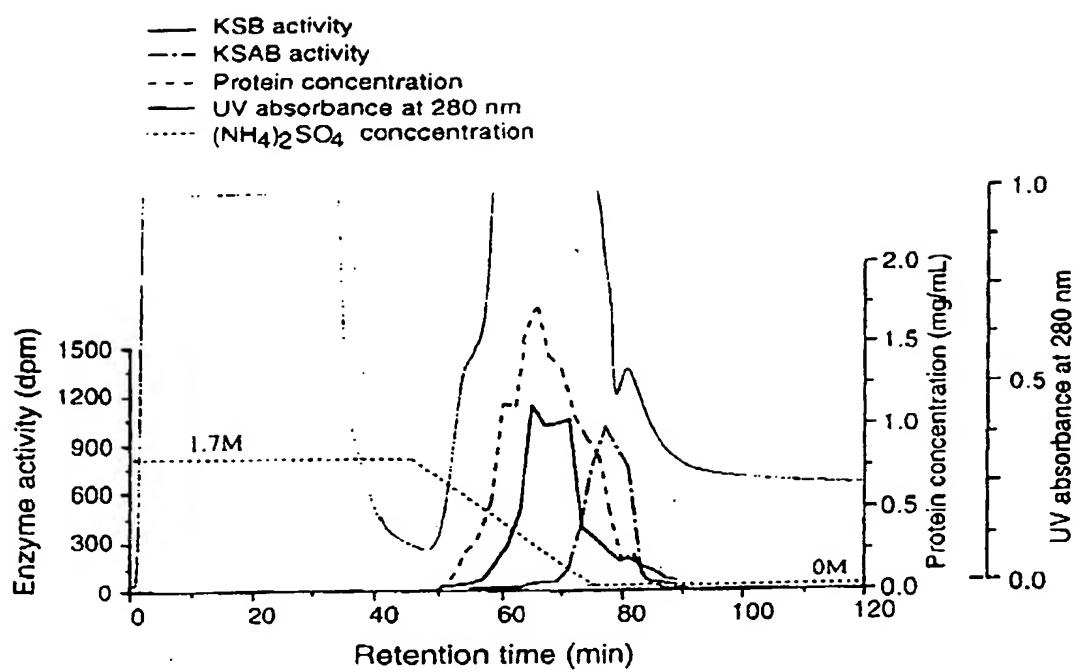


Fig. 2.

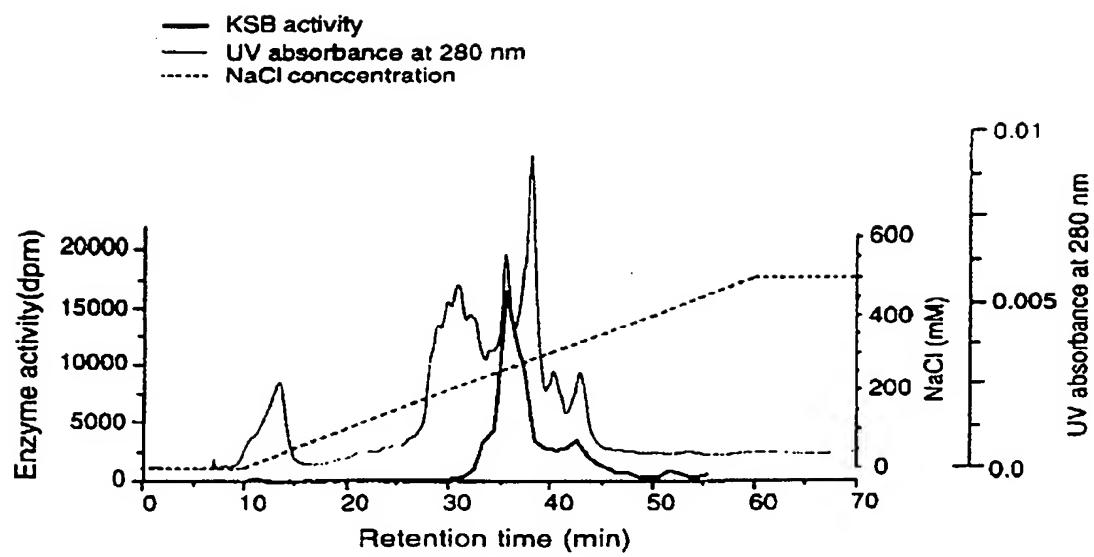


Fig. 3.

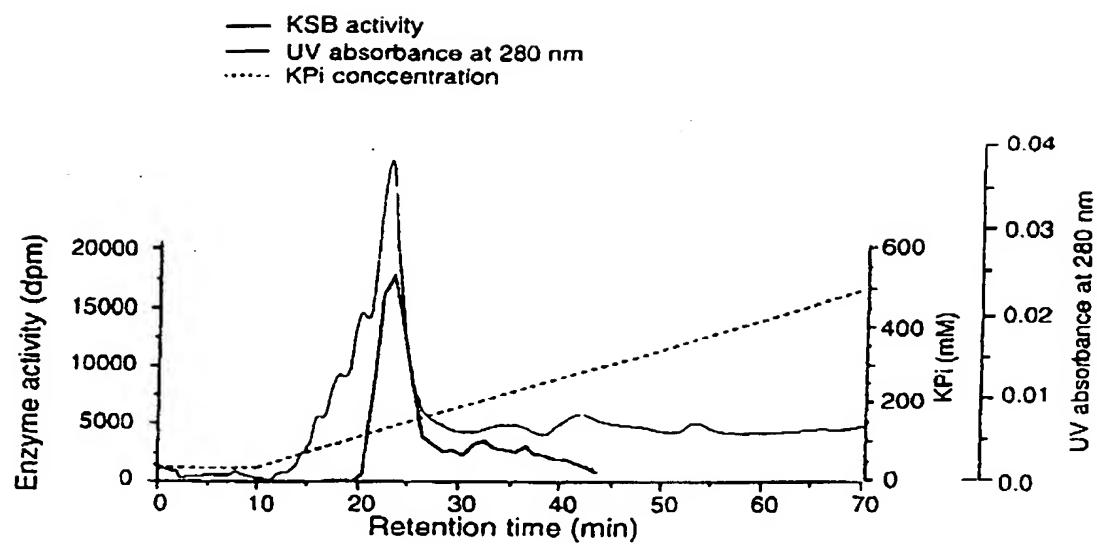


Fig. 4.

